Inherent problems with randomized clinical trials with observational/no treatment arms

Clifford J. Buckley, MD, Robert B. Rutherford, MD, Edward B. Diethrich, MD, and **Shirley D. Buckley, MSN, RN, CS,** *Temple, Tex*

Randomized clinical trials (RCTs) offering an observation/no treatment (OBS/NoRx) arm as control and which are focused on the management of a condition with potentially life-threatening consequences, however small the risk, often experience a significant rate of crossover to treatment by those randomized to the OBS/NoRx arm. Results of these trials when analyzed on intent-to-treat basis often fail to resolve the issue at which they were directed. The authors have observed this in trials of abdominal aortic aneurysms with this design and use these to exemplify the dilemmas RCTs of such design create, with crossovers ranging from 27% to over 60% (EVAR II, UKSAT, ADAM, PIVOTAL). Results of these trials are frequently used as level I medical evidence and their potential impact on clinical decision making and reimbursement can be quite significant and long-lasting. Recommendations regarding trial end points and suggestions to mitigate the high crossover effect are offered. It may be that some clinical conditions dealing with potentially life-threatening problems should not be studied in randomized prospective clinical trials containing an OBS/NoRx arm. (J Vasc Surg 2010;52:237-41.)

Prospective randomized clinical trials (RCTs) are generally considered a "gold standard" and are offered as level I medical evidence. By design, some RCTs have investigational arms offering observation or no treatment (OBS/NoRx). Unlike drug trials, RCTs investigating a device or procedure cannot utilize placebos and randomization to OBS/NoRx is transparent to the patient subject. Our experience in enrolling patients in RCTs with this design reflects an inherent problem with such RCTs in that a significant number of patients who initially accept randomization to OBS/NoRx ultimately cross over to a treatment arm. This is particularly true when the RCT is focused on a clinical condition that may be perceived by the patient, or family and friends, as life-threatening, such as abdominal aortic aneurysms (AAAs). The frequency with which patients cross over may either confound the outcomes of these trials and/or undermine the acceptance of conclusions based on an intent-to-treat analysis, which should not be used in this situation. Intent-to-treat data analytic strategy was developed for drug trials in which some patients dropped out (after 10-12 weeks) before receiving full treatment. To determine whether the full treatment worked, you could just use the subjects who completed treatment when analyzing outcome data for these studies. Our experiences with patients enrolled in a current AAA RCT exemplify this, but the extent of this problem is also apparent from reviewing other AAA RCTs with OBS/NoRx arms. These observations will be used to characterize the nature and extent of this problem.

From the Division of Vascular Surgery, Scott and White Hospital.
Competition of interest: none.
Correspondence: Clifford J. Buckley, MD, Division of Vascular Surgery, Scott and White Hospital, 2401 S. 31st Street, Temple, TX 76508 (e-mail: cbuckley@swmail.sw.org).
The editors and reviewers of this article have no relevant financial relationships to disclose per the JVS policy that requires reviewers to decline review of any manuscript for which they may have a competition of interest.
0741-5214/$36.00
Published by Elsevier Inc. on behalf of the Society for Vascular Surgery.
doi:10.1016/j.jvs.2010.02.255

OBSERVATIONS

We recently enrolled 37 patients into a "small" AAA (4- to 5-cm-diameter) RCT in which endovascular aneurysm repair (EVAR) is being compared with an OBS/NoRx arm featuring periodic ultrasound surveillance (US/S). During the first 36 months of the study, over 50% of subjects (10 of the 19) we enrolled in the OBS/NoRx arm crossed over to treatment. Although the reasons given for crossing over seemed insignificant to us, they clearly reflected the patients' growing concerns over not being treated for a condition, which, they felt, or had come to feel, deserved treatment. The reasons for requesting crossover were typically related by the patient as abdominal pain and/or tenderness but were frequently associated with comments made by relatives or friends or information gleaned from the Internet. This particular RCT's protocol recognized four acceptable reasons for crossing over from periodic US/S to EVAR: (1) AAA diameter reaching the threshold chosen for intervention; (2) a rate of increase in AAA diameter exceeding protocol interval growth limits; (3) the patient's AAA becoming symptomatic; and (4) patient requesting treatment. The last two reasons were invoked in 8 of 10 of our patients' crossing over from the observational

arm to EVAR. The responses of these patients ultimately evolved into a familiar pattern, as related below.

As part of the informed consent, patients were counseled at the outset to report any and all discomfort or pain related to the back, abdomen, or groin. This was understandably required because of the possibility of AAA expansion and rupture, albeit these were relatively small AAAs. Patient anxiety regarding no treatment for their AAA commonly appeared to increase during their periodic ultrasound surveillance, heightening with each US/S visit. It seemed to us that insignificant back pain, or anything causing abdominal pain or discomfort, seemed capable of triggering patients' concerns and prompted calls to the principle investigator's or study coordinator's offices between scheduled US/S visits. Visits to the emergency department or their primary physician's office were frequently interspersed between these calls. One patient called almost weekly seeking reinforcement regarding the safety of continued observation without treatment for his AAA. These calls usually related "new information" the patient had obtained on the Internet or from friends or relatives questioning the lack of treatment for an AAA that is known to be dangerous if not treated. The well-meaning concerns of family or friends regarding the risk of AAA rupture often magnified the patient's anxiety. The frequency of patient calls and office visits outside those scheduled, exerts considerable pressure on the private investigator and study coordinating staff. The alternative of dismissing repeated calls about apparently insignificant "symptoms" pose a potential litigation risk should the patient's AAA rupture while under observation. The ultimate outcome was often a demand to cross over to AAA repair. The national media coverage leading up to the passage of the Screening Abdominal Aortic Aneurysms Very Efficiently (SAAAVE) Act, with its discussions of the risks of AAA rupture, may have fueled patient concerns even though a main thrust of this Act is to use US/S to detect and follow smaller AAAs.[1]

Other AAA RCTs of similar design. An analysis of three other RCTs focused on the management of AAAs is included here to demonstrate the effect of patient crossover from observation to treatment in AAA trials and its impact on trial outcomes and conclusions based on intent-to-treat analysis. High crossover rates were encountered in both the United Kingdom Small Aneurysm Trial [2-4] and Aneurysm Detection and Management Veterans Affairs Cooperative (ADAM) Trial [5.]

In the United Kingdom Small Aneurysm Trial,[4] 1090 patients ages 60 to 76 with aneurysms 4 to 5.5 cm in diameter were randomized to either open repair (OR) or US/S. Of the 563 patients randomized to OR, 34 expired while awaiting repair, including 9 who ruptured their aneurysm. Of the 527 patients randomized to surveillance, 327 (62%) crossed over to OR by the end of the 4.6-year mean follow-up period. Less than 38% of patients in the observation arm actually completed their randomized assignment. The majority crossed over because of aneurysm-related "symptoms" rather than increase in aneurysm size. Additionally, 63 patients crossed over because of individual preference or protocol deviation. At last report,[6] only 218 remained in the observational arm of the trial (19% of the total). Some interpreted the trial as showing it was not a matter of whether or not to repair such AAAs, but when!

The ADAM trial compared the results of immediate OR with US/S in 1136 veterans (ages 50-79) with 4 to 5.5 cm in diameter AAAs. In this RCT, 569 were randomized to OR and 567 to US/S. During the 4.8 years mean follow-up, 61.6% of those randomized to US/S crossed over to OR, whereas only 7.3% of those randomized to OR did not undergo OR. Reasons for crossover included "symptoms" related to the AAA; aneurysm diameter exceeding the upper threshold of 5.5 cm; increase in AAA diameter of more than 1 cm per year; need for intervention for iliac or thoracic aortic aneurysm or patient request. The majority of patients crossed over because either their aneurysm became "symptomatic" or patient request. The risk of aneurysm rupture in all patients in this trial was estimated to be 0.6% per year. Interestingly, two patients awaiting OR and 11 patients in the surveillance arm ruptured their AAAs during the period of the trial.

In the United Kingdom Endovascular Aneurysm Repair Trial II (EVAR-II),[7] patients over the age of 60 with an AAA of greater than or equal to 5.5 cm diameter and who were judged "unfit" (too high risk) for open surgical repair were randomized to either EVAR or OBS/NoRx. Of three hundred thirty-eight patients participating in this study, 166 were randomized to EVAR and 172 to OBS/NoRx. In this trial, 47 patients (~27%) crossed over from the OBS/NoRx arm to AAA repair. In the EVAR arm, 14 of the patients (~8%) died before receiving EVAR, six from AAA rupture and eight from other causes, presumably related to comorbidities in these high-risk patients. Ultimately, 197of all 338 randomized patients (~58%) received some form of aneurysm repair (150 randomized and surviving to undergo EVAR and 47 crossovers). The AAA repair received was EVAR in 181 and OR in 16. Only 125 of 172 patients randomized to OBS/NoRx (73%, but only 37% of the total in the trial) remained in that arm during the duration of the trial. The 47 patients who crossed over to EVAR from OBS/NoRx did so for a variety of reasons, the published list including aneurysm-related "symptoms", rapid rate of expansion, or "patient or physician preference" (Table). Although full details were not given, only five of the 47 patients, those whose AAAs had increased significantly in size, had what might be conceded as nonsubjective reasons to cross over to repair. This significant crossover rate, of course, was not reflected in the intent-to-treat analysis, which showed no significant difference in overall mortality between those assigned to EVAR (~45%) and those assigned to OBS/NoRx (~40%). The published report does state that a per-protocol analysis was done and showed no significant difference between these two arms in terms of all-cause mortality ($P = .07$) and aneurysm-related death ($P = .43$), but some suspected that these analyses did not take into account those (8%) who did not receive EVAR as assigned because they died while awaiting repair. However, even if these 14 patients are excluded, in addition to includ-

Table. Analysis of number of patients crossing-over to treatment from observation/no treatment in the randomized clinical trials cited

Study	*ADAM n = 1136*	*EVAR-II n = 338*	*UK small AAA n = 1090*	*Our experience n = 37*
Treatment arm	569	166	563	18
OBS/NoRx	567	172	527	19
No. of crossover to repair	349	47	327	10
Reasons for crossover:				
AAA size	10	0	0	0
AAA diameter increase	0	5	0	2
Patient request or off protocol	31	14	63	2
"Symptomatic" or increasing anxiety	297	26	247	6

AAA, Abdominal aortic aneurysm; *ADAM,* Aneurysm Detection and Management Veterans Affairs Cooperative; *EVAR,* endovascular aneurysm repair; *OBS/NoRx,* offering observation or no treatment.

ing the crossovers, and, if one also makes allowances for the 16 patients receiving OR, not EVAR, the mortality of those actually receiving EVAR (71/181 39%) is still not statistically significant compared with those remaining under observation (57/125 46%) ($P = .29$). Parenthetically, 21 of the latter patients died from AAA rupture. This trial is an example of one in which the apparent effect of crossovers undermined its acceptance. The differences in mortality rate percentages being so suggestive. Even though per-protocol analyses did not significantly favor EVAR, a lack of acceptance was also likely fueled by differences in outcome between crossovers and those assigned to receive EVAR. The 2% initial and 23% late mortality rates for those who crossed over and received EVAR compared with the 9% initial and 40% late mortality of those originally assigned to EVAR were statistically significantly different (initial $P = .005$, late $P = .0008$).[7,8] Cronenwett, in an invited critique of these trials for Lancet commented to the effect that the EVAR II trial proved that patient/physician choice is superior to randomization.[9] The latter differences also became the focus of subsequent US reports showing much lower procedural mortalities for EVAR in high-risk cohorts from the Lifeline registry (2.9%)[10] and NISQIP data from the Veterans Administration (3.1%).[11] These seemed more commensurate with the 2% mortality of crossovers than the 9% mortality of those assigned to EVAR. These publications indicated that US investigators, some performers of EVAR, were unwilling to accept the results and conclusions of this "flawed trial", in spite of the crossovers making a statistically significant difference in mortality in a per-protocol analysis.

DISCUSSION

Patients with a vascular disorder are generally willing to participate in an RCT addressing its management for altruistic reasons in that the information derived from the trial may help others even if not themselves. Most trials are based on equipoise and the fact that the patient may have been made aware when two or more treatment modalities are being compared it is understood that no one therapy has been established as "the best of those being compared." Where some form of therapy is being offered to those enrolled in all arms of the trial, the patient may accept randomization as long as no bias regarding which treatment is superior exists. In contrast, those RCTs dealing with a perceived life-threatening condition (ie, AAA) and, which offer an OBS/NoRX arm present difficulties in patients' accepting randomized assignment to OBS/NoRx and staying with it.

Many patients being evaluated for AAA already possess some information regarding the indications for AAA repair, the different procedural options involved and have some appreciation for both the risks and benefits associated with these options. Patient bias was encountered in many, if not most, of the trials for FDA approval of the various aortic stent grafts. Although it was originally intended that EVAR be randomly compared with OR, randomization to OR was often not accepted by patients who really "sought" EVAR, possibly aware of the "hype" associated with the stent grafts that emphasized EVAR's potential of lower morbidity and faster return to normal activities. Because of patient bias, ultimately the "controls" in these FDA trials had to be made up of those with unsuitable anatomy for endovascular repair. Indeed, it is not uncommon today for AAA patients to present themselves not only expecting EVAR but even identifying the particular stent graft they have heard has the best treatment record. Various forms of "hearsay", including Internet sites accessed by the patient or their family or friends, or anecdotal comments by friends or family, all play a role in producing this crossover phenomenon.

When initially screened, patients with little previous information may readily accept randomization to an OBS/NoRx arm because of the perception that observation without treatment carries an extremely low risk for adverse outcome. However, as these patients continue to participate in the trial, they naturally investigate their condition further and, if they obtain "information" from family or friends or the Internet, which leads them to believe their condition is potentially life-threatening if untreated, they become increasingly concerned of the advisability of continuing in an OBS/NoRx arm. The end result often is that rather than openly challenge information regarding risk received from the private investigator or study coordinators, they may dwell on or magnify "symptoms", which

could be construed as indicating a worsening of their AAA. Repeated calls and other expressions of these growing concerns exert increasing pressure on the investigator or trial coordinating staff who is obliged to approve their eventual request to cross over to treatment.

Many patients, or those close to them, perceive, or come to perceive, an AAA as a "ticking time bomb" and wish to be relieved of its threat. This may be reflected in the subsequent improvement beyond an initial period of 3 to 6 months in quality of life observed in those undergoing AAA repair vs those who did not.[12,13] The observation group of an AAA trial are reminded of the potential rupture risk every time they are required to return for US/S. This may be likened to a "Sword of Damocles effect", ie, a perceived risk, the removal of which brings relief. Despite medical knowledge of the relationship of AAA size to rupture risk and its use as positive reinforcement by healthcare professionals participating in small AAA trials, the patient's perception of the risk for AAA rupture ultimately will prevail, and since most clinical trials with OBS/NoRx arms appropriately offer the opportunity for crossover to treatment related to changes in clinical status or patient desire, this option ends up being taken by a substantial portion of the patients over time with a significant impact on the outcome of the trial or the acceptance of conclusions drawn therefrom.

The end point of most AAA RCTs is mortality, but it can be argued that the primary objective when treating patients with AAAs is not to prolong life, for many have significant life-threatening comorbidities, but to prevent AAA rupture or, at least, death from that event. Although the three cited trials failed to show an overall mortality advantage for treatment, their data appear to support the conclusion that both OR and EVAR prevent AAA rupture. In the EVAR-II Trial, only one of 213 patients (0.5%) whose AAA was repaired had AAA rupture, while 8 AAA ruptures occurred in the 125 surveillance patients (6.4%). This represents a significant difference ($P = .002$ using Fisher exact test). In the UK Small Aneurysm Study, three of 890 repaired patients (0.03%) had AAA rupture, while 26 ruptures (17 under surveillance plus 9 waiting repair) occurred in 125 patients not receiving a repair (20.8%). This also represents a significant difference ($P > .0001$ using Fisher exact test). US/S of patients with AAAs in itself obviously offers no protection from rupture, although it may detect "too-rapid" growth or growth reaching an accepted interventional threshold diameter and thus dictate the need for repair. However, compliance with US/S surveillance of AAAs may be better in some (eg, European) countries than in North America[14] and, in fact, it has shown to be poor in one US study, even when surveillance is free. Valentine reported 32% noncompliance in a United States small aneurysm surveillance trial with 27% failing to keep any scheduled appointments and 5% lost after one to two follow-up visits. Four of these noncompliant patients had documented AAA rupture.[15]

We believe our concerns about AAA trials of this design are appropriate. The question remains, is there a solution to this dilemma? Some of the following might be considered in future AAA trials at least: (1) Randomization in RCTs of this design might be weighted more heavily toward the OBS/NoRx arm (ie, 2:1), to compensate for inevitable high crossover rate, especially if intent-to-treat analysis is the primary method used. (2) Drop subjects who cross over in either direction. While this has the benefit of maintaining the randomized trial, it could be argued that the sample is now biased if significantly more subjects cross over from the NoRx to Rx arms. (3) Per-protocol analysis should always be reported, in addition, comparing those actually receiving treatment versus those remaining in the OBS/NoRx arm. (4) RCTs of this design should not be conducted in health care systems in which waiting times for elective admission are characteristically long and the time between randomization and receiving the assigned treatment cannot be shortened for trial enrollees (eg, a mean of 57 days for AAAs of mean diameter of 6.7 cm in the EVAR I trial). (5) Trials of this design should be restricted to small AAAs (ie, those where the risk of rupture under surveillance is acceptably low (ie, ~1%-3%/year). (6) Should treatment with any agent currently undergoing trial (eg, roxithromycin, doxycycline, statins, angiotensin-converting enzyme inhibitors, beta receptor blockers, or protease inhibitors) be shown to reduce AAA growth, they could be included in the surveillance arms of AAA trials in the future, as best medical therapy and thus help to reduce the crossover rate.[16] (7) Although both all-cause mortality and AAA-related death should continue to be reported (the latter along with the autopsy rate for deaths occurring in the period beyond 30 days), AAA rupture should also be at least a secondary end point for such RCTs. Without such changes, conclusions derived from AAA trials with OBS/NoRx arms are likely to have unacceptably high crossover rate and their results and conclusions are unlikely to gain acceptance, even though they are considered to constitute level I medical evidence.

It may well be that certain vascular diseases, which have a perceived life-threatening outcome, should not be evaluated using RCTs containing an OBS/NoRx arm. Human nature coupled with the patient's desire for survival predisposes the patient to want therapy for a perceived life-threatening condition.

CONCLUSIONS

Based on an experience in a current AAA RCT and our analysis of other AAA RCTs that offer OBS/NoRx arms, we conclude that this trial design fails to take human nature into account and results in significant crossover rates to the treatment arm in AAA trials. This can obfuscate the comparison of management options using intent-to-treat analysis. RCTs of this design or their conclusions have usually led to significant controversy, in large part, related to the numbers of subjects who cross over to the treatment arm. Nevertheless, their potential impact on clinical decision making and reimbursement can be quite significant. This inherent weakness has commonly resulted in these trials not resolving to the satisfaction of many, if not most, clinicians the issue at which they were directed. We have offered some

possible solutions which, although they have yet to be put to the test in mitigating this dilemma, may at least offer additional helpful perspective in viewing the outcomes of these trials.

REFERENCES

1. http://www.vascularweb.org/professionals/Government_Relations/National_Aneurysm_Alliance_SAAAVE_Bill/ 2006.
2. The UK Small Aneurysm Trial Participants. Mortality results for randomized controlled trial of early elective surgery or ultrasonographic surveillance for small abdominal aortic aneurysms. Lancet 1998;352:1649-55.
3. The UK Small Aneurysm Trial Participants. The UK Small Aneurysm Trial: design, methods, progress. Eur J Vasc Endovasc Surg 1995;9:42-8.
4. The UK Small Aneurysm Trial Participants. Long-term outcomes of immediate repair compared with surveillance of small abdominal aortic aneurysms. N Eng J Med 2002;346:1445-52.
5. Lederle FA, Wilson SE, Johnson GR, Reinke DB, Littooy FN, Archer CW. Immediate repair compared with surveillance of small abdominal aortic aneurysms. N Engl J Med 2002;346:1437-44.
6. Powell JT, Brown LC, Forbes JF, Fowkes FG, Greenhalgh RM, Ruckley CV. Final 12-year follow-up of surgery versus surveillance in the UK Small Aneurysm Trial. Br J Surg 2007;94:702-8.
7. EVAR Trial Participants. Endovascular aneurysm repair and outcome in patients unfit for open repair of abdominal aortic aneurysm (EVAR trial 2): randomized controlled trial. Lancet 2005;365:2187.
8. Bush RL, Mureebe L, Bohannon WT, Rutherford RB. Research review. The impact of recent European trials on abdominal aortic aneurysm repair: Is a paradigm shift warranted? J Surg Res 2008;148:264-71.
9. Cronenwett JL. Endovascular aneurysm repair: important mid-term results. Lancet 2005;365:2156-8.
10. Siccard GA, Zwolak RM, Sidawy AN, White RA, Siami FS. Society for Vascular Surgery Outcomes Committee. Endovascular abdominal aortic aneurysm repair: long-term outcome measures in patients at high-risk for open surgery. J Vasc Surg 2006;44;229-36.
11. Bush RL, Johnson ML, Hedayati N, Henderson WG, Lin PH, Lumsden AB. Performance of endovascular aortic aneurysm repair in high-risk patients: results from the Veterans Affairs National Surgical Quality Improvement Program. J Vasc Surg 2007;45:227-33.
12. Lederle FA, Johnson GR, Wilson SE, Archer CW, Ballard DJ, Littooy FN, et al; Aneurysm Detection and Management Veterans Affairs Cooperative Study Group. Quality of life, impotence and activity level in a randomized trial of immediate repair versus surveillance of small abdominal aortic aneurysms. J Vasc Surg 2003;38:745-52.
13. Perkins JM, Magee TR, Hands LG, Collin J, Galland RB, Morris PJ. Prospective evaluation of quality of life after conventional abdominal aortic aneurysm surgery. Eur J Vasc Endovasc Surg 1998;16:203-7.
14. Cronenwett JL, Johnston KW. The United Kingdom Small Aneurysm Trial: Implications for surgical treatment of abdominal aortic aneurysms. J Vasc Surg 1999;29:191-4.
15. Valentine RJ, Decaprio JD, Castillo JM, Modrall JG, Jackson MR, Clagett GP. Watchful waiting in cases of small abdominal aortic aneurysms-appropriate for all patients? J Vasc Surg 2000;32:441-8.
16. Baxter BT, Terrin MC, Dalman RL. Medical management of small abdominal aortic aneurysms. Circulation 2008;117:1883-9.

Submitted Mar 25, 2009; accepted Feb 18, 2010.